Metadata Profile for the National Digital Catalog

March 17, 2008; version 1.0

Introduction

Metadata is documentation about a resource that can be provided at several levels of detail. This profile document addresses discovery-level metadata used to facilitate searching for geological and geophysical data contained in collections across the U.S. Geological Survey, State Geological Surveys, the Minerals Management Service, and the Bureau of Land Management. The National Digital Catalog of Geological and Geophysical Data and its needs for discovering physical data resources and how to access them forms the intended purpose of a National Digital Catalog as outlined in the Implementation. The standard is:

- Derived from ISO 19115 Geographic Information – Metadata
- Derived as a profile within the process outlined in ISO 19106 Geographic Information – Profiles
- Managed by the National Catalog Committee representing State Geological Surveys and U. S. Department of the Interior partners

Scope

The National Digital Catalog consists of three levels of "data objects" that are used as a reference for the metadata profile.

- The National Catalog – The entire data resource that will be found in the catalog can be considered on one level as a single, geospatially oriented database. In actual fact, it will be made up of many individual databases provided by data collection owners but will be aggregated together with a set of simple standards to provide a conglomerate data resource.
- Data Collections – Individual collections reported in Phase 1 of the National Catalog effort are themselves distinct, geospatially oriented (in most cases) databases of geological and geophysical resources.
- Sample-level Metadata – Each collection contains an inventory of individual samples. Phase 2 of the National Catalog will focus on site-specific (containing a sample collection geospatial footprint) records of geological and geophysical data with a standard set of metadata elements.

Profile

The National Catalog Committee has elected to make use of the ISO 19115 Geospatial Information – Metadata standard and the associated ISO 19139 Geospatial Information – Metadata—XML Schema Implementation as maintained by the Open Geospatial Consortium (http://schemas.opengis.net/iso/19139/). These standards and the

accompanying profile that specifies how the elements are contributed to the National Catalog provide a basis for maximum application flexibility within a set of adopted standards.

The discovery-level metadata profile makes use of a Phase 1 National Catalog survey process to identify the basic information on data collections that correspond to high-level fields such as "originator" and "distributor" of data along with general methods of accessing underlying physical data for a given collection. The profile provides guidance on the additional metadata elements being collected during Phase 2 of the National Catalog in one of three ways:

- XML template (see file download)
- CSV spreadsheet template (see file download)
- XML Web service based on the template

Certain data fields will accept longer text strings (e.g., Abstract). In the XML template, these are shown as "CDATA" sections that can accept HTML or other rich text formatting. If using the CSV template, these attributes should not contain any special characters and should be qualified with double quotes.

Other data fields will accept multiple values (marked with "N" under Number of Occurrences in the Metadata Elements table). In the XML template, these fields contain multiple possible values; use as many as necessary. In the CSV template, these attributes can be supplied as a comma-delimited list (e.g., a list of URLs in the onlineResource field). These fields should also be qualified with double quotes.

Normative References

The concept and some overall structural elements for this document were inspired by the UK GEMINI Standard: A Geo-spatial Metadata Interoperability Initiative. (http://www.gigateway.org.uk/metadata/pdf/UK_GEMINI_v1.pdf)

Implementation Plan for the National Geological and Geophysical Data Preservation Program (http://datapreservation.usgs.gov/docs/2006DataPreservation.pdf)

ISO 10115: Geographic Information – Metadata (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020)

ISO 19139: Geographic Information – Metadata—XML Schema Implementation (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32557)

ISO 19106: Geographic Information – Profiles (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26011)

Metadata Elements

The following table presents the attributes represented in the templates and provides guidance to be followed in providing site-specific records for the collection inventory phase of the National Catalog project.

The table shows:

- Each data element (represented as XML attributes or as columns in the spreadsheet template)
- Corresponding attribute names from the ISO 19115 standard (in italics where applicable)
- A description or definition of the element
- Whether the element is mandatory (M), in which case the value must be provided, or optional (O), its multiplicity – single valued (1) or multi-valued (N) [Multi-valued elements must have replicates of XML attributes as indicated in the template or be placed in a comma-separated list in the spreadsheet, retaining the quoted column definition as provided in the template.]

Element Name	**Definition**	**Obligation**	**Number of Occurrences**
CollectionID	The unique collection ID assigned during Phase 1 of the National Catalog effort that identified distinct collections via an online survey. Collection IDs are available online from http://datapreservation.usgs.gov/ and are used to derive additional sample-level metadata.	M	1
Title *MD_DataIdentification.Citation*	The human-readable title for the individual record that will be used in any listing or search result. Only one official title will be used and may be truncated in certain displays.	M	1

Element Name	Definition	Obligation	Number of Occurrences
AlternateTitle *CI_Citation.alternateTitle*	Collection owners may elect to provide additional title identifiers for individual records for further identification or use by other Web service interfaces. The AlternateTitle field may include either textual titles or specific sample IDs used by the collection.	O	N
Abstract MD_Identification.abstract	The human-readable description of the individual record used to help determine the nature of the underlying physical data resource. Due to the general nature of the catalog, a fair amount of information about the data resource may need to be captured into this one general element.	M	1
DataType *MD_Keywords*	A controlled vocabulary of data types has been derived from Appendix 2 in the NGGDPP Implementation Plan and is provided in a table below. Each reported physical data record must fall into one of the types from the list. A collection inventory report provided via one of the templates can contain records from many different types. *The data types provided in this context will be considered "discipline-specific" keywords in the ISO 19115 standard.*	M	1

Element Name	Definition	Obligation	Number of Occurrences
SupplementalInformation *MD_DataIdentification.supplementalInformation*	This standard field will be used to provide specific information on how to access the physical data represented by the metadata record. This may be general for the entire collection (e.g., a URL to another Web site) or a specific reference to an online resource like an ordering system with a specific ID.	M	N
Coordinates	In order to facilitate rapid and standardized incorporation into a National Catalog, the provided records must be a simple Geographic or latitude and longitude coordinate for the sample collection point. In the templates, these should be listed as longitude,latitude in the same field. *Regular bounding box coordinates can be derived from this simple point in compliance with the ISO 19115 standard.*	M	1
AlternateGeometry	The underlying collection resource may use an alternate method of storing a geospatial footprint. If so, this text field should be used to describe the authoritative source for geographic location and how the simple coordinates were derived.	O	1
OnlineResource *CI_OnlineResource*	One or more URL pointers to textual information about the specific record.	O	N
BrowseGraphic *MD_BrowseGraphic*	One or more URL pointers to images representing the specific record.	O	N

Element Name	Definition	Obligation	Number of Occurrences
Date *EX_TemporalExtent.extent*	If a meaningful date within the geosciences domain can be attached to the record (e.g., a collection date), it can be supplied here. Either date may be to any degree of precision, or may be left blank to indicate uncertainty. Examples are 20010101, 1939-1945, -20030331, 2000-.	O	N
DatasetReferenceDate *CI_Citation.date*	A reference date indicating currency of the underlying data record. In many cases, this may be the date the metadata record was assembled for the National Catalog. Proper date formats are defined in ISO 8601.	M	1
VerticalExtent *EX_VerticalExtent*	If applicable, vertical extent information can be provided for the specific record. For example, vertical depth information can be very useful for rock core samples. Specification of extent can contain three elements: minimum value, maximum value, and unit of measure. For the purpose of the National Catalog, these elements will be collected as 2 or 3 values representing the UnitOfMeasure and MaximumValue with the possibile addition of MinimumValue (e.g., m,35.4,0 for a rock core measured at 35.4 total meters).	O	1

Data Type Classification

The Implementation Plan for the NGGDPP identifies the basic data types with which the program concerns itself. For the purposes of Phase 2 of the National Digital Catalog, the National Catalog Committee developed a list of data types that will be addressed as sample-level metadata for the National Catalog. The following standard values should be used in classifying data type for individual sample records.

- Auger Sample
- Fluid Sample
- Geochemical Sample
- Hand Sample
- Ice Core
- Paleontological Sample
- Rock Core
- Rock Cuttings
- Sediment Core
- Sidewall Core
- Thin Section
- Type Stratigraphic Section

The other data types identified in the Implementation Plan are all valid data that will ultimately be cataloged and made accessible for discovery services in the National Catalog. However, the National Catalog effort will first address the broad range of geological and geophysical data represented in the list that can be connected to a specific geographic point.